

## 2

### Explaining Cellular Phenomena through Mechanisms

I do not in the least mean by this that our faith in mechanistic methods and conceptions is shaken. It is by following precisely these methods and conceptions that observation and experiment are every day enlarging our knowledge of colloidal systems, lifeless and living. Who will set a limit to their future progress? But I am not speaking of tomorrow but of today; and the mechanist should not deceive himself in regard to the magnitude of the task that still lies before him. Perhaps, indeed, a day may come (and here I use the words of Professor Troland) when we may be able 'to show how in accordance with recognized principles of physics a complex of specific, autocatalytic, colloidal particles in the germ cell can engineer the construction of a vertebrate organism'; but assuredly that day is not yet within sight of our most powerful telescopes. Shall we then join hands with the neovitalists in referring the unifying and regulatory principle to the operation of an unknown power, a directive force, an archæus, an entelechy, or a soul? Yes, if we are ready to abandon the problem and have done with it once and for all. No, a thousand times, if we hope really to advance our understanding of the living organism.

(Wilson, 1923, p. 46)

The focus of this book is creation of cell biology in the mid-twentieth century as a distinct field of biology devoted to discovering and understanding the mechanisms that account for the ability of cells to live. The notion of mechanism has a long history in philosophy (a brief sketch follows), but it was largely eclipsed in twentieth-century philosophy of science, which emphasized laws and, not coincidentally, physics rather than biology. Mechanisms have begun again to be the focus of discussion in very recent philosophy of science. I will devote a major portion of this chapter to discussing the central features of this newly emerging conception of mechanism and mechanistic explanation. While many of the philosophers advancing this notion of mechanism have

been principally concerned with biology, the notion of mechanism on which consensus has settled is insufficiently biological. Accordingly, a further step in my discussion will be to develop a more biologically adequate notion of mechanism, one that is itself inspired by the conception of a cell as the basic living unit. Because actual scientific investigation, such as that pursued in cell biology, is not concerned with the abstract character of a mechanism, but with investigating the details of actual mechanisms, I will conclude by examining how mechanisms are studied, and the challenges cell biology faced in discovering the mechanisms in cells that account for the activities of life.

## 1. HISTORICAL CONCEPTIONS OF MECHANISM

In western thought, the notion of mechanism originated with the design by the ancient Greeks of machines, such as the wedge, to perform work. Greek thought, however, usually contrasted mechanics with nature – machines facilitated accomplishing work that opposed natural forces. This is clear in the pseudo-Aristotelian text *Mechanica*:

Nature often operates contrary to human interest, for she always follows the same course without deviation, whereas human interest is always changing. When, therefore, we have to do something contrary to nature, the difficulty of it causes us perplexity and art has to be called to our aid. The kind of art which helps us in such perplexities we call Mechanical Skill. (*Mechanica* 847a14f)

Accordingly, nature itself was not conceived as operating mechanically. Aristotelian philosophy in particular advanced an anti-mechanistic conception of nature. It emphasized *telos*, the end state to be achieved by entities of nature, and the form, which resided in bodies and determined their nature and what they did. Thus, Aristotle stressed the distinctive forms (souls) of living things and the activities they made possible (nutrition and reproduction for plants; also sensation and locomotion for animals; and all of these plus reason for humans).

The Aristotelian perspective was still dominant when, in the seventeenth century, numerous natural philosophers rebelled and developed a program of explaining natural phenomena mechanically. Unlike the Greek mechanics, the seventeenth-century mechanists identified natural phenomena as themselves mechanical and offered mechanical explanations of naturally occurring activities.

Although Aristotelian mechanics set nature in opposition to mechanism, the Greek atomist tradition advanced a view closer in spirit to the

later mechanical philosophy. Such philosophers as Leucippus (ca. 480–ca. 420 BCE) and Democritus (ca. 470–ca. 380 BCE) were called *atomists* because they sought to explain phenomena in nature by appealing to their constituent elementary particles, or atoms. They characterized these particles in terms of their shape and size, and then appealed to these properties to explain the characteristics of the macroscopic objects they comprised. Thus, Democritus proposed that hot bodies were hot because they were composed of elements of fire, which were small and round, whereas cold bodies were cold because they were composed of larger particles with sharp points. The early modern mechanists followed the atomists in appealing to the shapes of the hypothetical minute components of material objects to explain the properties of macroscopic objects.

One of the most prominent of the early modern mechanists was Galileo Galilei (1564–1642). He is celebrated for developing a mechanics of simple moving bodies, an account he extended to explain celestial phenomena such as the movement of the earth around the sun. He also developed hypothetical mechanisms to account for mundane phenomena, and here the inspiration of the atomists is most clear. For example, he offered the following explanation of the power of heat to liquefy bodies:

The extremely fine particles of fire, penetrating the slender pores of the metal . . . fill the small intervening vacua, and . . . set free these small particles from the attraction which these same vacua exert upon them and which prevents their separation. Thus the particles are able to move freely so that the mass becomes fluid and remains so long as the particles of fire remain inside; but if they depart and leave the former vacua, then the original attraction returns and the parts are again cemented together. (Galilei, 1638/1914, p. 19)

Rene Descartes (1596–1650) provided perhaps the fullest development of the mechanical philosophy in the early modern period. As he said in *Principia*, “I have described this earth and indeed the whole universe as if it were a machine: I have considered only the various shapes and movements of its parts” (Descartes, *Principia* IV, section 188). In appealing to the motions of the parts as well as their shapes, Descartes added to the resources of the ancient atomists. A key component of his mechanistic conception was that the movement of one object would cause movement in another. In particular, because Descartes did not allow for empty space, any movement of one object required the movement of other objects.

Having emphasized the change caused by the contact of one body with another, Descartes, in contrast to both the ancient mechanics and to the Aristotelians who had seen mechanics as opposed to nature, argued for treating the

natural world as a mechanism and all events in it as the result of the operation of a mechanism. In so doing he denied any ontological difference between natural phenomena and the workings of human made mechanisms:

I do not recognize any difference between artifacts and natural bodies except that the operations of artifacts are for the most part performed by mechanisms which are large enough to be easily perceivable by the senses – as indeed must be the case if they are to be capable of being manufactured by human beings. The effects produced by nature, by contrast, almost always depend on structures which are so minute that they completely elude our senses. (Descartes, *Principia* IV, section 203)

Descartes' conception of mechanism extended as well to the biological world, including the functioning of human bodies. In this he appealed frequently to metaphors with human-made machines, using them to suggest explanations of features of biological systems that initially appear to set them apart. For example, to explain the ability of animals to initiate motion, he appealed to such artifacts as clocks and mills that succeeded in moving on their own. Although William Harvey had already offered his own mechanical pump model for the circulation of blood, Descartes proposed to explain circulation as resulting from heating that caused the expansion of droplets of blood, which then forced their way through the arteries. He further appealed to the hydraulically moved statuary in the Royal Gardens to provide a model of the ability of the nervous system to transmit sensory and motor signals from the brain. He proposed that in animals the brain directed activity by altering the flow of very fine matter, the animal spirits, through the nerves (see Figure 2.1).

A generation after Descartes, the British chemist Robert Boyle (1627–91) coined the term *mechanical philosophy*. Boyle directed part of his empirical work to developing the air pump as a device for creating vacuums. In allowing for a vacuum, that is, space unoccupied by objects, Boyle developed a different perspective on a mechanistic conception of nature than Descartes had endorsed. Furthermore, his interest in the air pump was not just that of an engineer: Boyle sought to explain the ability of air to exert pressure, and proposed a mechanical model of air molecules as small springs that could be compressed and later expand. As exemplified in his account of air, Boyle's general strategy was to appeal to the shape and motion of hypothetical small particles to explain the properties of different chemical substances.

Although the notion of mechanism was a major feature of early modern philosophy, another tradition developed in the same period ultimately superseded it in philosophical accounts of science. Instead of explaining the



Figure 2.1. Descartes' representation of the nerve transmission in reflex action. Animal spirits traverse the path from the sense organs to the brain and back, resulting in motion. Reprinted from René Descartes (1664), *L'Homme*. Paris: Charles Angot.

activities of the natural world as workings of a mechanism, this tradition placed principal emphasis on *laws* of nature. Ronald Giere (1999), among others, has proposed that this view was a product of the theological perspective of certain early modern philosophers – those who conceived of the principles governing the natural world on the model of civil laws, albeit with God, not humans, as the lawgiver. Appeal to a lawgiver continued as a feature of this tradition, but what gave the law-based tradition traction was that scientists learned to describe regularities in the universe economically and precisely in laws that were appealed to in explaining particular phenomena. The perspective is clearly manifest in Newton, who posited forces in nature to explain phenomena from the terrestrial to the celestial. His well-known three laws of motion are an exemplar of this approach, but he pursued the approach more generally. To explain the reflection of light, for example, he posited a repulsive force between a light ray emitted and its source as soon as a ray left the area of attraction. Vaporization was similarly attributed to a repulsive force. Moreover, Newton suggested a systematic relationship between

repulsive and attractive forces in general: “As in Algebra, where affirmative Quantities vanish and cease, there negative ones begin; so in Mechanics, where Attraction ceases, there a repulsive Virtue ought to succeed” (*Opticks*, Query 31).

What disturbed many mechanists about Newton’s approach was that he envisaged forces as operating over a distance, whereas the mechanical philosophy, especially as developed by Descartes, insisted on contact in any causal interaction. Newton himself, in the preface to *Principia*, tried to downplay this difference, referring to forces as mechanical principles:

I wish we could derive the rest of the phenomena of Nature by the same kind of reasoning from mechanical principles, for I am induced by many reasons to suspect that they may all depend upon certain forces by which the particles of bodies, by some causes hitherto unknown, are each mutually impelled towards one another, and cohere in regular figures, or are repelled and recede from one another. These forces being unknown, philosophers have hitherto attempted the search of Nature in vain; but I hope the principles here laid down will afford some light either to this or some truer method of philosophy. (1687, Preface to the First Edition)

As Boas (1952) discussed in her classic history of the mechanical philosophy, two schools of Newtonians developed. One school accepted forces operating without contact while the other insisted on maintaining some version of mechanical contact as involved in all actions of forces:

One, comprising those whom we call Newtonians today, accepted the concept of action at a distance; rejected mechanical explanations, denying their necessity; converted the ether into magnetic and electric fluids where necessary; and generally followed the theories enunciated in the *Principia* rather than the more speculative hypotheses of the *Opticks*. But there were other physicists of the eighteenth century who followed the more mechanical aspects of Newton’s theory and used his ether to form a highly complex system. (Boas, pp. 519–20)

## 2. TWENTIETH-CENTURY CONCEPTIONS OF MECHANISM

The dominant twentieth-century philosophy of science, developed by philosophers and scientists who characterized themselves as *logical empiricists*, adopted a non-mechanical emphasis on explanation in terms of laws. Part of the reason for this was the logical empiricists’ generally Humean, anti-metaphysical perspective, which rejected positing hidden causal relations between events. According to Hume, we observe sequences of events but do not observe anything connecting one event to the next. Hume favored a

minimalist concept of cause and effect, in which an event that regularly precedes another type of event can be denoted as its cause; Hume thought no further explanation should be sought. Being generally skeptical of the human capacity to satisfactorily address metaphysical issues, the logical empiricists went a step further and tried to avoid reference to causes by focusing instead on the linguistic statement of laws. They developed a conception of explanation, known as the *deductive-nomological model* of explanation, that involved derivation of an observation statement from statements of laws along with statements specifying initial conditions (Hempel & Oppenheim, 1948).

In contrast to the logical empiricists, Wesley Salmon (1984) developed an account of scientific explanation that made the notion of cause itself central. For Salmon, the way to answer the question of why something happened was to identify what caused it. Salmon tried to formulate a non-problematic account of what a cause is, but his proposals have in turn been the focus of critical objections.<sup>1</sup> I will not pursue this debate because my interest in Salmon is not in his analysis of causation, but his characterization of his account as “causal-mechanical.” Salmon was one of the first twentieth-century philosophers of science to revive the interest in mechanistic explanation. It is not clear, however, what Salmon intended by invoking the term *mechanical*. As Glennan (2002) noted, Salmon never explicated the notion of a mechanism, settling instead for such comments as “explanations reveal the mechanisms, causal or other, that produce the facts we are trying to explain” (Salmon, 1989, p. 121).

Before advancing an account of what a mechanism is, though, let me note an important respect in which Salmon’s account offers a significant advance over seventeenth- and eighteenth-century mechanical philosophy. Whereas the early modern mechanists allowed only such properties as the shape and motion of particles to figure in accounts of mechanisms, Salmon (1984, p. 241) broadened the category to include such constructs as force fields: “We have to change our mechanistic view from the crude atomism that recognizes only the motions of material particles in the void to a conception that admits such non-material entities as fields, but for all of that, it is still a mechanistic world view. Materialism is untenable, but the mechanical philosophy, I believe, remains viable.” This expansion repairs the breach in the mechanical philosophy introduced by Newton’s appeal to forces. Mechanical interactions no longer require contact between causes and their effects, but can accommodate more distal actions such as gravitational and magnetic attraction. (This expansion of the

<sup>1</sup> See Salmon (1984) for his initial proposal. For a critical response, see Kitcher (1989). Salmon (1994) proposed an alternative account, which in turn has been criticized by Dowe (1995).

conception of mechanical activity had in practice long become commonplace. For example, nineteenth-century mechanistic accounts of physiological processes incorporated energy-intensive operations such as chemical bonding and catalysis.)

### 3. CURRENT CONCEPTIONS OF MECHANISM

With increased attention given to the biological sciences in the last two decades, a number of philosophers of science have begun attending to mechanistic explanation. They have addressed the absence of an appropriate framework by offering initial proposals that overlap in some important respects but also vary in terminology, scope, and emphasis (see, for example, Bechtel & Richardson, 1993; Glennan, 1996; Machamer, Darden, & Craver, 2000).<sup>2</sup> I first provide a basic characterization of mechanisms as found in nature and then elaborate it into a framework for mechanistic explanation:

A mechanism is a structure performing a function in virtue of its components parts, component operations, and their organization. The orchestrated functioning of the mechanism is responsible for one or more phenomena.

Moreover:

- The component parts of the mechanism are those that figure in producing a phenomenon of interest.

<sup>2</sup> With Robert Richardson, I characterized mechanistic explanations as accounting “for the behavior of a system in terms of the functions performed by the parts and the interactions of these parts . . . . A mechanistic explanation identifies these parts and their organization, showing how the behavior of the machine is a consequence of the parts and their organization” (Bechtel & Richardson, 1993, p. 17). Whereas we focused on the “functions” (operations) that parts perform, Glennan emphasized the properties of parts in stating what he originally (1996) called laws and now (2002) calls “invariant change-relating generalizations.” These are instantiated in “interactions” in which “a change in a property of one part brings about a change in a property of another part” (2002, p. S344). Machamer, Darden, and Craver characterize mechanisms as “entities and activities organized such that they are productive of regular changes from start or set-up to finish or termination conditions” (2000, p. 3), emphasizing the distinct metaphysical status of “entities” (parts) and “activities” (operations). Tabery (2004) has proposed a partial synthesis in which activity and property changes are seen as complementary. I discuss the use of the term *operation* rather than *activity* in note 7 below. Finally, Machamer et al. (p. 3) include a characterization of mechanisms as “productive of regular changes from start or set-up to finish or termination conditions.” I am concerned that such an emphasis helps to retain a focus on linear processes whereas mechanisms, when they are embedded in larger mechanisms, are continuously responsive to conditions in the larger mechanism, for tractability scientists tend to focus on the conditions in which an operation is elicited and on its contribution to the behavior of the overall mechanism. However, they often have to counter this analytical perspective to appreciate the dynamics at work in the system.



- Each component operation involves at least one component part. Typically there is an active part that initiates or maintains the operation (and may be changed by it) and at least one passive part that is changed by the operation. The change may be to the location or other properties of a part, or it may transform it into another kind of part.
- Mechanisms may involve multiple levels of organization.
- Operations can be organized simply by temporal sequence, but those in biological mechanisms tend to exhibit more complex forms of organization.
- Mechanisms can be dynamic and can change both ontogenetically and phylogenetically.

Several features of this characterization of a mechanism require elaboration.

### *Mechanisms Explain Phenomena*

The conception of a mechanism is intimately tied to the context of explanation: Mechanisms are identified in terms of a phenomenon for which explanation is sought. In logical empiricist philosophy of science there was a tendency to construe explanations as explaining observation statements, which were taken to be theory-neutral characterizations of events. This view became problematic with the contention, stemming from Hanson (1948) and Kuhn (1962/1970), that observations are theory-laden in that what scientists observe is influenced by the theories they hold. This seemed to threaten circularity with theories being tested by observations that are already shaped by the theories being tested. Although there are simpler ways of showing that this circularity is not vicious,<sup>3</sup> Bogen and Woodward challenged the very idea that observations are what scientists explain. They contrasted observations and phenomena. Observations provide data but – except when the data does not turn out as expected and the investigator seeks to explain why – it is not data that scientists explain. Rather, they explain *phenomena* – occurrences in the world about which data can be procured.<sup>4</sup> Although there can be singular phenomena (the big bang or the birth of a particular organism), the

<sup>3</sup> For example, it is sufficient to note that even if the way in which we observe something is affected by our theories and other beliefs, this does not entail that we can observe whatever the theory predicts. Holding a theory that grass turns red in normal daylight will not make it the case that I will see red grass.

<sup>4</sup> Some purported phenomena do not exist. Bogen and Woodward cited the example of N-rays, a phenomenon that French physicists at the beginning of the twentieth century posited on the basis of several observations; ultimately, more careful data collection demonstrated that such a phenomenon did not exist.

phenomena of interest in science can be captured in generalizations involving, for example, a functional relationship between variables, or the fact that a certain kind of event regularly occurs only if a certain other type of event has just occurred. Data play an important role in identifying and providing evidence for phenomena, but it is the phenomena so identified that are the objects of explanation.

Bogen and Woodward offered as examples of phenomena “weak neutral currents, the decay of the proton and chunking and recency effects in human memory” (1988, p. 306). In biology, DNA replication or alcoholic fermentation would be comparable examples. It is often possible to characterize phenomena quantitatively. Bogen and Woodward consider the example of lead melting at 327° C. Galileo established that the distance traveled by an object falling freely near the surface of the earth is sixteen times the square of the number of seconds it falls. A biological example of a quantitative phenomenon is that the maximum number of molecules of adenosine triphosphate (ATP) formed in normal cells via oxidative phosphorylation per oxygen molecule consumed is three. Phenomena may also be characterized with different degrees of specificity. An individual scientist might, for example, identify the phenomenon for her investigation as the synthesis of a particular protein in a specific type of cell occurring in a particular species living under specified conditions. The author of a review paper might address the phenomenon of synthesis of that protein in a variety of cell types and species. At the most general level, a textbook author might write a few pages simply on “protein synthesis.”

Identifying and characterizing phenomena is a challenging scientific activity that consumes considerable resources of time, money, and ingenuity. A purported phenomenon must be shown to be genuine and its generality determined. Some purported phenomena do not stand up under scrutiny and must be discarded. Although I will emphasize the importance of a specification of the phenomenon for the development of mechanistic explanations, it is important to note at the outset that scientists often revise their characterizations of phenomena in the course of investigating the mechanisms they take to be responsible for them. In *Discovering Complexity*, Richardson and I referred to such revisions as *reconstituting the phenomenon* and we offered the example that, in the course of investigating the mechanism of gene expression, researchers repeatedly revised the conception of what genes code for. In the 1860s Gregor Mendel spoke of factors for traits. In the 1910s Thomas Hunt Morgan and his collaborators sought to localize genes for such traits as eye color, but in the 1940s Beadle and Tatum’s inquiry with mutations in *Neurospora* led them to link genes instead to individual enzymes.

If the revision to the conception of the phenomenon is so major that little of what the mechanism was thought to be doing is still recognized as occurring, the work toward articulating the mechanism for the phenomenon as originally characterized may prove to have been in vain. Most often, though, the change in the characterization of the phenomenon takes the form of a revision, not a wholesale replacement of the old conception, and the changes in the account of the mechanism are accordingly more restricted. For example, early investigators construed fermentation as a catabolic activity breaking down sugar and yielding alcohol (with heat as a by-product). Once researchers recognized that the energy released was captured in high-energy phosphate bonds that were used as energy sources for other cell activities, the conception of the phenomenon to be explained was revised. It was now a mechanism for converting the energy of foodstuffs into a form useful in such other activities as cell division. Yet, because the catabolic breakdown of sugar to alcohol remains part of this process, much that had been learned about the mechanism of fermentation still applied after the phenomenon had been reconceptualized.

It bears emphasizing that the project of providing explanations, including mechanistic explanations, starts with the identification of a phenomenon. This is where the functioning structure gets determined, constraining what will count as a successful identification of relevant parts and operations and their organization (Kauffman, 1971). If the operation of some entity does not contribute to the production of a given phenomenon, it is not part of the mechanism responsible for that phenomenon. On this construal, different mechanisms may be instantiated in the same substance in the same spatial-temporal region and may share many component parts and operations. What unites one set of parts and operations into a given mechanism is their organization and their orchestrated functioning in producing a particular phenomenon.

I will conclude this discussion of phenomena by introducing an illustration I will continue to use to characterize various features of a mechanism. After the investigations of Harvey, pumping of blood through the circulatory system was a well delineated phenomenon. Although we now take the phenomenon to be obvious, until Harvey established the more general phenomenon of the circulation of the blood, the phenomenon of pumping blood was not recognized. Rather than conceiving of circulation, investigators assumed that both the arteries and veins transported material to the bodily tissues and that this phenomenon was readily accounted for as a result of newer material pushing older material along. Once Harvey established that the blood circulated, the need for a pump to move blood was recognized and the functioning

heart was identified as the mechanism responsible for this phenomenon.<sup>5</sup> The importance of specifying the phenomenon to be explained is illustrated with this example: Until the heart was recognized as performing the function of pumping blood, there was no interest in understanding the way in which this occurred. Moreover, hearts also do other things – they make sounds and someone might want to explain that phenomenon. That, however, is a different phenomenon that involves a different mechanism – a system of parts and operations that presumably shares some components with the mechanism for pumping blood but is not identical to it.

### *Component Parts and Component Operations*

The next aspect of mechanisms to emphasize is that mechanisms consists of component parts and component operations.<sup>6</sup> Figure 2.2 illustrates key components of the heart viewed as a mechanism for pumping blood. Component parts of the heart include the atria and ventricles, the valves between the atria and ventricles and between the ventricles and arteries, and the blood itself. Component operations include the contraction and relaxation of the atria and ventricles and the opening and closing of the valves. Blood is forced out of the atria and ventricles as they contract, and prevented from flowing back by the closing of the valves afterward.<sup>7</sup> The blood here is a part of the mechanism, but one that is operated on rather than itself performing operations (in the context of this phenomenon). Although in this case the parts that perform

<sup>5</sup> In many cases, however, the entity or system responsible for the phenomenon is not immediately obvious and must be discovered. For example, in Bechtel and Richardson (1993), we discussed the extended controversy in the nineteenth century of the locus of control for respiration until Pflüger (1875) established that it occurred within individual cells.

<sup>6</sup> When I am emphasizing the thing performing the operation, I use the term *part* or *component part* and when I am emphasizing what the part does, I speak of *operation* or *component operation*. I sometimes use the term *component* alone where it is not important to be specific, or to designate jointly the part and the operation it performs.

<sup>7</sup> Machamer, Darden, and Craver employ the term *activity* to draw attention to the fact that components of mechanisms are active. The mechanical philosophy of the seventeenth century departed from Aristotelian philosophy in construing elements of nature as passive, doing something only if acted upon. In some accounts, the giant clock-like machine of nature was wound up at the outset and motion in the world is the playing out of that initial state. The term *activity*, however, does not readily capture the fact that in most operations there is also something acted upon. This is the reason I have preferred the term *operation*. Typical of the operations I have in mind are the reactions of chemistry which prototypically involve a catalyst, a reactant, a product, and often a cofactor. In some reactions there is no need for a separate catalyst as the energetic factors are such that the reaction will occur spontaneously. In autocatalytic reactions, which are highly relevant in living systems, the product of the reaction also serves as a condition for more iterations of the reaction.

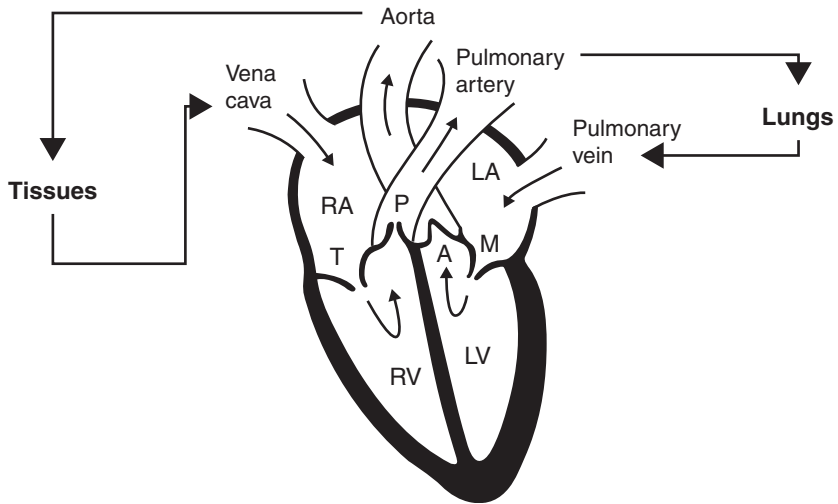


Figure 2.2. An example of a mechanism: the heart pumping blood. Labeled parts include RA: right atrium; LA: left atrium; RV: right ventricle; LV: left ventricle; T: tricuspid valve; M: mitral valve; P: pulmonary valve; A: aortic valve.

operations are separate from those operated on, in other cases the parts that perform operations may also be affected by the parts on which they operate (as I will discuss in Section 6, such feedback provides a prime way in which mechanisms can be self-regulating).

The component parts and operations of a mechanism do not present themselves to the scientist neatly distinguished and labeled as in a textbook. The investigations resulting in an understanding of the mechanism require *decomposing* it (taking it apart), conceptually if not physically. Corresponding to the division of parts and operations, there are two types of decomposition. What I refer to as *structural decomposition* decomposes a structure into component parts while *functional decomposition* decomposes the function into component operations. Though sometimes coordinated, it is not uncommon for these two types of decomposition to be pursued independently of each other, by scientists in different fields employing different tools. Often one decomposition proceeds more rapidly and successfully than the other, with considerable time elapsing before the slower inquiry catches up.

One example of structural decomposition is the discovery via anatomical dissection that (1) the body has a heart; and (2) the heart has four chambers (RA, LA, RV, and LV) and at least four valves (T, M, P, and A). Another example is the discovery via microscopy that (1) tissues are composed of cells; (2) cells contain a plasma membrane, nucleus, and cytoplasm; (3) cytoplasm

contains an aqueous cytosol and a variety of organelles such as mitochondria and the Golgi apparatus; (4) each organelle has an internal structure (description and further levels varying by organelle). An example of functional decomposition is the discovery via physiological investigation that the overall function of pumping blood includes multiple component operations of contraction and relaxation (at different times in different chambers) and of opening and closing (of valves). The parentheticals indicate that it is generally difficult to specify operations without some indication of the parts involved. It is useful to have a separate notion of functional decomposition, though, because progress in identifying operations often can proceed when there is minimal knowledge of some or all of the parts involved. For example, biochemists in the early twentieth century decomposed the overall function of cellular respiration into numerous biochemical reactions (operations) while, structurally, they had basic knowledge of the substrates and products (passive parts) but little more than invented names for the enzymes (active parts) that were assumed to catalyze the reactions.

Ultimately, the full characterization of a mechanism requires mapping the operations into which the overall function of the mechanism is decomposed onto the parts into which the structure is decomposed. I use the term *localization* for such mappings and discuss them further below. Uncovering the organization of the parts and operations at a given level – not merely identifying them – is also crucial. Such a combined perspective is often required to understand fully how the mechanism generates its behavior, because frequently the spatial layout of the parts enables or facilitates the temporal organization of their operations. As a practical matter, moreover, structure and function frequently provide critical insights into the other. Learning the structural character of a part can provide insight into how it carries out its operation. Understanding the operation that is being performed often provides clues as to what sort of part is responsible. As we will see, many of the major contributions of modern cell biology to understanding the mechanisms responsible for the phenomena exhibited in living cells involved the ability to relate various organelles to the physiological operations they perform and, at a lower level, certain components of the organelles with particular biochemical operations.

### *Organization and Orchestration*

A point just mentioned – that it is important to determine how parts and operations are organized – deserves elaboration. A mechanism is typically not just a collection of independent parts, each carrying out its operation in isolation. Rather, parts and operations are generally integrated into a cohesive,

functioning system. In the heart, the veins, atria, ventricles, and arteries must be appropriately spatially related to each other. Further, the valves must be located at the right place and oriented so as to prevent blood from flowing backwards through the system. This exemplifies the fact that the parts of a mechanism are organized. Its operations also are organized. Each may occur in its turn – a simple temporal ordering – or there may be a good deal of overlap, interdependency, or other complications. As the timing becomes more complex, especially in mechanisms with many parts, we further say that the operations are *orchestrated* so as to produce the phenomenon of interest. For example, the phenomenon of the heart pumping blood depends on the spatial organization of the component parts, the temporal organization of the component operations (movements), and the fine-grained spatiotemporal orchestration of the moving parts in real time.

There are several reasons organization is so important. If one part is to operate on the product produced by the operation of another part, it needs to have reliable access to that product. One way to ensure this is for the two parts to be spatially contiguous. Another is to provide a mode of transport between them. The assembly lines used in human manufacturing adopt this mode of organization – components are brought to the line and added to the emerging product in sequence. Organization in biological systems, however, is seldom as simple as the sequential arrangement used in assembly lines. One of the key features of organization in biological mechanisms is the incorporation of feedback and other kinds of control systems that allow the behavior of some components of the mechanism to be regulated by other components of the mechanism. The role of more complex modes of organization is a major feature that renders the biological conception of mechanism different from that which suffices for non-biological mechanisms, as discussed further in Section 6.

#### 4. REPRESENTING AND REASONING ABOUT MECHANISMS

Mechanisms are real systems in nature, which led Salmon (1984) to identify his causal-mechanical approach to explanation as *ontic*, as it appeals to the actual mechanism in nature. He contrasted it with an *epistemic* conception of explanation that appeals to laws and derivations from laws, which are clearly products of mental activities. Salmon's insight is important, but the ontic/epistemic distinction does not properly capture it. He is right that in mechanistic explanation a scientist appeals to causal relations and mechanisms operative in nature, which are taken to generate or produce the

phenomena being explained. However, it is important to note that offering an explanation is still an epistemic activity and that the mechanism in nature does not directly perform any explanatory work.<sup>8</sup>

There are several ways to appreciate the epistemic character of mechanistic explanation. First, the mechanisms operative in our cells were operative long before cell biologists discovered and invoked them to explain cellular phenomena. The mechanisms are not themselves the explanations; it is the scientist's discovery and rendering of aspects of the mechanism that produces what counts as an explanation. Second, the difference between the mechanism and the mechanistic explanation is particularly obvious when considering incorrect mechanistic explanations – in such a case a scientist has still appealed to a mechanism, but not one operative in nature. Such a mechanism exists only in the representation offered by the scientist. It is thus the mechanism as represented, not the mechanism itself, that figures in explanation. (It is also the phenomenon as represented that scientists seek to explain.) Thus, scientists offer a mechanistic explanation by identifying and representing parts and operations regarded as key to producing the phenomenon and showing how, appropriately organized, they can do so.<sup>9</sup>

Mechanisms can be represented either in linguistic descriptions or in diagrams. Philosophical accounts of science have tended to privilege linguistic representations and regard diagrams as at best crutches for following the linguistic argument. When one considers the actual practice of scientists in reading papers, however, the tables seem to be turned. It is common for readers to scan the abstract and then jump to key figures. To the extent that crutches are involved, the figure captions that provide commentary on the figures play this role. Consider a paper in which a mechanistic explanation is proposed. The diagram provides a vehicle for representing the complex interactions among operations, while the commentary can only characterize these one at a time. The text of the paper then provides yet further commentary: about how the mechanism is expected to operate (introduction), how evidence as to its operation was procured (methods), what evidence was advanced (results), and the interpretation of how these results bear on the proposed mechanism

<sup>8</sup> This point is eloquently stated by Christian de Duve (1984, p. 18) at the outset of his masterful *Tour of the Living Cell*: “Every object, every site, every happening, every process, every mechanism that will be pointed out as though it were there to be seen is actually a product of individual human brains mulling and churning over collections of images and sets of figures, themselves the products of recordings made by intricate instruments on biological materials subject to complex experimental manipulations.”

<sup>9</sup> I thank Cory Wright for impressing on me that a mechanism in nature does not itself explain anything.



(discussion). The detailed commentary is important, but it is the diagram that represents the mechanism. As just one example of the saliency of diagrams, Christian de Duve, whose role in discovering the lysosome will be discussed in detail in Chapter 5, recalls that his discovery of the lysosome was sparked by an unexpected failure in his biochemical investigation of a liver enzyme. “By some fortunate coincidence, my recent readings had included [two 1946 papers by Claude and] I immediately recalled Claude’s diagrams showing the agglutination at pH 5 of both large and small granules, and concluded that our enzyme was likely to be firmly attached to some kind of subcellular structure” (de Duve, 1969, p. 5).

The importance scientists place on diagrams should lead us to question whether they are in fact superfluous. Are there reasons a scientist might prefer to represent certain information diagrammatically rather than propositionally? More importantly, are there different processes of reasoning with diagrams than with propositions such that an account of science that focused only on logical inference would fail to capture an important aspect of explanatory reasoning?

The motivation for using diagrams to represent mechanisms is obvious. Unlike linguistic representations (except those found in signed languages), diagrams make use of space to convey information. As the heart example revealed, spatial layout and organization is often critical to the operation of a mechanism itself. As in a factory, different operations occur at different locations. Sometimes this serves to keep operations separate from one another and sometimes it serves to place operations in association with one another. These spatial relationships can be readily shown in a diagram. Even when information about the specific spatial layout is lacking or not significant, one can use space in the diagram to relate or separate operations conceptually. Moreover, diagrams can take advantage of dimensions other than space that visual processing can access, including color and shape.<sup>10</sup>

Time is at least as important as space to the operation of a mechanism – one operation proceeds, follows, overlaps, or is simultaneous with another operation. This can be captured by using one of the spatial dimensions in a diagram to convey temporal order. This of course presents a problem: Most diagrams are two dimensional and this leaves just one dimension for everything other than time. One solution – as exemplified in the heart diagram – is

<sup>10</sup> These can either be iconic – representing the actual color or shape of the parts of a mechanism, or they can be symbolic. fMRI diagrams of brain activity are a well-known example of the symbolic use of color – colors scaled from hot to cold are used to represent such things as strong to weak activations or high to low statistical significance of the increase of the activation above some baseline.

to make strategic use of arrows to represent temporal relations, leaving both dimensions free to represent the mechanism's spatial or similarity relations. Another solution is to use techniques for projecting three dimensions onto a two-dimensional plane.

Whether the temporal order of operations is represented by means of a spatial dimension or by arrows, a diagram has clear advantages over linguistic description. The most obvious advantage – that all parts and operations are available for inspection simultaneously – probably is the weakest one. Due to processing limitations, people can take in only one or a few parts of the diagram at a time. Nonetheless, more so than when reading text, they have the freedom to move around it in any number of ways; and as the diagram becomes more familiar, more of it can be taken in at one time. A stronger advantage is that diagrams offer relatively direct, iconic resources for representation that can be invaluable. For example, it is immediately apparent in the heart diagram that blood is being pumped simultaneously from the two atrial chambers to the two ventricles and that these two parallel operations are in a sequential relationship to two other parallel operations (pumping from the two ventricular chambers).

The value of consulting a diagram in this way is even more apparent in mechanisms with feedback loops, through which an operation that is conceptually downstream (closer to producing what is taken to be the product of the mechanism) has effects that alter the execution of operations earlier in the stream at subsequent time steps. Multiple examples can be found within the cellular respiration. As biochemists discovered in the 1930s, it is composed of three connected submechanisms (as illustrated in the next chapter in Figure 3.16). When these are further unpacked, they are seen to involve coordinated biochemical operations, including feedback operations. Figure 2.3 shows an important feedback loop that operates at the interface between the first two submechanisms – glycolysis and the citric acid cycle. The diagram aids understanding by spatially laying out the parts of the system (compounds such as pyruvate) and by using the vertical dimension as well as arrows to indicate the sequence of operations (solid arrows for the reactions and a dotted arrow for the feedback loop).

An important principle recognized by cognitive scientists engaged in modeling reasoning computationally is that it is essential to coordinate the modes of representation and procedures of inference. If diagrams are an important vehicle for representing mechanisms, then it is necessary to consider how people reason about diagrams. Philosophers since Aristotle have often assumed that procedures of logic, especially natural deduction, describe our reasoning. But logic operates only on linguistic representations, so if scientists reason

### Phosphoenolpyruvate

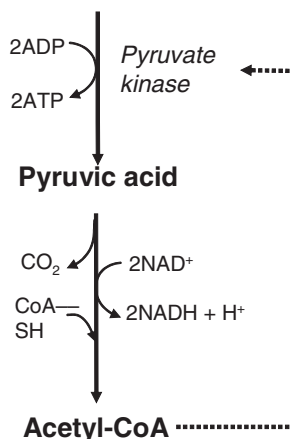


Figure 2.3. Feedback loop in the linkage between glycolysis and the citric acid cycle. In the final reaction of glycolysis, phosphoenolpyruvate produces pyruvic acid. Pyruvic acid then produces acetyl-CoA, some amount of which is needed to continuously replenish the citric acid cycle (not shown). If more acetyl-CoA is produced than can be used in the citric acid cycle, it accumulates and feeds back (dotted arrow) to inhibit pyruvate kinase, the enzyme responsible for the first step in the reaction. This in turn will stop glucose from entering the glycolytic pathway.

with diagrams, the operations of reasoning must be different. To understand how scientists reason with diagrams it is helpful to keep in focus the fact that mechanisms generate the phenomenon in virtue of their component parts performing their operations in a coordinated manner. The kind of reasoning that is needed is reasoning that captures the actual operation of the mechanism, including both the operations the components are performing and the way these operations relate to one another.

One limitation of diagrams when it comes to understanding mechanisms is that they are static. Even if they incorporate arrows to characterize the dynamics of the mechanism, the diagram itself doesn't do anything. Thus, it cannot capture the relation of the operation of the parts to the behavior of the whole mechanism. Accordingly, the connection together must be provided by the cognitive agent. The cognizer must imagine the different operations being performed, thereby turning a static representation into something dynamic.<sup>11</sup>

<sup>11</sup> Animated diagrams relieve people of this difficult task and are often far more instructive to novices. Thomas M. Terry of University of Connecticut has produced some excellent ones that make clear how the many operations in cellular metabolism are related. He has them posted at <http://www.sp.uconn.edu/~terry/images/anim/ETS.html>. Another good site for such diagrams,

Mary Hegarty (1992) called the activity of inferring “the state of one component of the system given information about the states of the other system components, and the relations between the components” *mental animation* and emphasized its importance to the activities of designing, troubleshooting, and operating mechanical devices (p. 1084). Obtaining reaction time and eye movement data while people solved problems about relatively simple pulley systems, she investigated the extent to which inference processes are isomorphic to the operation of the physical systems. One way they were not isomorphic is that the participants made inferences about different components of the system (i.e., individual pulleys) separately and sequentially even though in the physical system the components operated simultaneously. The participants found it considerably harder, however, to make inferences that required them to reason backward through the system rather than forward, suggesting that they animated the system sequentially from what they represented as the first operation, in this respect preserving isomorphism with the actual system.

Accepting the claim that people, including scientists, understand diagrams of mechanisms by animating them, a natural follow-up question concerns how they do this. A plausible initial proposal is that they create and transform an image of the mechanism so as to represent the different components each carrying out their operations. In perception we have experience of parts of the system changing over time, and so the proposal is that in imagination we animate these components by invoking the same processes that would arise if we were to watch an animated diagram. This proposal needs to be construed carefully, as a potential misunderstanding looms. Reference to a mental image should not be construed as reference to a mental object such as a picture in the head. Recent cognitive neuroscience research indicates that when people form images they utilize many of the same neural resources that they do in perception (Farah, 1988; Kosslyn, 1994).<sup>12</sup> Thus, what occurs in the head in forming an image is activity comparable to that which would occur when seeing an actual image. Barsalou (1999) speaks of this neural activity as a *perceptual symbol*.

which also provides links to Terry’s diagrams, is <http://www.people.virginia.edu/~rjh9u/atpyield.html>.

<sup>12</sup> Within cognitive science there has been a heated controversy over whether the representations formed in the cognitive system are really image-like (Kosslyn, 1981; Kosslyn, 1994; Pylyshyn, 1981; Pylyshyn, 2003). This discussion can remain neutral on this issue since the fundamental issue is not how the cognitive system encodes its representations but what it represents something as. The visual system represents objects as extended in space and changing through time. What is important here is that scientists can represent mechanisms mentally in much the way that they represent diagrams that they encounter (albeit with less detail than when actually looking at the diagram).

Thinking with perceptual symbols then involves the brain initiating sequences of operations that correspond to what it would undergo if confronted with actual input from visual objects behaving in a particular manner. Barsalou refers to this as *simulation*. Moreover, simulation is not restricted to repeating those sequences of neural processes that occurred in previous experience. Just as we can imagine objects that we have never seen by recombining components of things we have seen, we can imagine sequences of changes that depart from those that we have actually encountered.

Although humans are relatively good at forming and manipulating images of rather simple systems, if what we are imagining is the working of a rather complex mechanism that has multiple components interacting with and changing each other, we often go astray. We fail to keep track of all the changes that would occur in other components of the system in response to the changes we do imagine. Thus, the usefulness of mental animation for understanding a mechanism does reach a limit. Ordinary people may simply stop trying at this point, but scientists and engineers often find it important to do better and hence have created tools that supplement human abilities to imagine a system in action. One tool involves building a scale model (or otherwise simplified version) of a system and using it to determine how the actual system would behave. The behavior of the scale model *simulates* that of the actual system. For example, the behavior of objects in wind tunnels can be used to simulate phenomena involving turbulence in natural environments. If instead an investigator can devise equations that accurately characterize the changes in a system over time, the investigator can often determine how the system will behave by solving the equations without actually building a scale model. In this case the simulation is done with a mathematical model rather than a physical model. The advent of the computer provided both a means of solving the equations of a mathematical model and an additional means of simulating systems. Higher level computer languages are designed to represent complex structures and their interactions, and by using these resources, one can often create a computer simulation of the interactions in a complex system (Jonker, Treur, & Wijngaards, 2002).

These different modes of simulating a system all provide an important advantage when a mechanism is complex with multiple operations occurring simultaneously – they do not lose track of some of the interactions as a human imagining the operation of the mechanism often does. But even when it is a human who is doing the imagining, what he or she is doing can also be characterized as simulating the mechanism.

Although a mechanism can be represented by means of a diagram, it can also be described linguistically. Is there any fundamental difference between

linguistic and diagrammatic representations? Larkin and Simon (1987) considered diagrams and linguistic representations that are informationally equivalent and analyzed how they can nonetheless differ with respect to ease of search, pattern recognition, and the inference procedures that can be applied to them. In part these differences stem from the fact that information that may be only implicit in a linguistic representation may be made explicit, and hence easier to invoke in reasoning, in a diagram (Larkin & Simon, 1987, p. 65).<sup>13</sup> More recently, Stenning and Lemon (2001) suggested that diagrams are more constrained in expressive power than propositions and accordingly are more tractable. They also argued that the advantage provided by these constraints is dependent upon the subject supplying an interpretation that makes them available.

## 5. LEVELS OF ORGANIZATION AND REDUCTION

The part–whole relationship between a mechanism’s component parts and its structure can be understood as falling within the type of hierarchical, mereological framework that systematic biologists and others have long used to bring orderliness to types of entities at different levels. The relationship between a mechanism’s component operations and its overall function have roughly the same character, though less attention has been paid to systematizing this kind of relationship. What is important here is that both kinds of components (the parts and their operations) can be regarded as occupying a lower level than the mechanism itself (a structure with a function). Because of this difference in levels, mechanistic explanations are commonly characterized as *reductionistic*.<sup>14</sup> The notion of reduction that arises with mechanistic explanation, however, is very different from that which has figured either in popular discussions or in recent philosophy of science, and its consequences are quite different. In these discussions, appeals to lower levels are thought to deny the efficacy of higher levels. While the functioning of a mechanism

<sup>13</sup> Larkin and Simon commented, “In the representations we call diagrammatic, information is organized by location, and often much of the information needed to make an inference is present and explicit at a single location. In addition, cues to the next logical step in the problem may be present at an adjacent location. Therefore problem solving can proceed through a smooth traversal of the diagram, and may require very little search or computation of elements that had been implicit” (1987, p. 65).

<sup>14</sup> Wimsatt (1976) was one of the first philosophers to recognize the relation between mechanism and what most scientists mean by *reduction*: “At least in biology, most scientists see their works as explaining types of phenomena by discovering mechanisms, rather than explaining theories by deriving them or reducing them to other theories, and *this* is seen as reduction, or as integrally tied to it” (p. 671).

depends upon its constitution, it also depends on its context, including its incorporation within systems at yet higher levels of organization. Mechanistic reductionism neither denies the importance of context or of higher levels of organization nor appeals exclusively to the components of a mechanism in explaining what the mechanism does. The appeal to components in fact serves a very restricted purpose of explaining how, in a given context, the mechanism is able to generate a particular phenomenon.

Before explicating further what mechanistic reduction involves, I need to clarify the notion of level. The notion of level is widespread in both philosophical and scientific discussions (see Churchland & Sejnowski, 1992) and it is often assumed that levels cut across all of nature so that there are levels of subatomic particles, atoms, molecules, and so forth. Extending this to the living world, there are levels of cellular organelles, cells, tissues, organs, organisms, societies, etc. There are problems in fitting such a conception of levels together with the scientific practice of explaining phenomena because often the crucial operations in nature cross these levels – electrons interact with molecules, ions with membranes, and single-celled organisms with organisms containing multiple organs. Accordingly, it seems wise to abandon the attempt to demarcate levels that transect nature; rather, I restrict the identification of levels to local contexts in which mechanistic explanations are offered for particular phenomena.

The notion of level enters into discussions of mechanisms in virtue of the fact that a mechanistic explanation decomposes a mechanism into its component parts and operations. Thus, an investigator begins with a phenomenon – for example, the phenomenon of organisms taking in oxygen and releasing carbon dioxide and water. In the attempt to explain this phenomenon the investigator decomposes the structure (the organism's body) into component parts, decomposes its function (respiration) into component operations, and determines how they are organized and orchestrated to produce the phenomenon. With respect to this phenomenon, the operative parts of the mechanism – the lungs, blood, tissues, etc. – constitute entities at a lower level of organization than the respiring organism. Another investigator may be interested in how tissues perform their role in this mechanism and decompose them into cells and then, within cells, into different organelles involved in cellular respiration. These components are then at a lower level than those identified in the first decomposition. In principle this process can continue indefinitely, but in practice it stops after two or three rounds of decomposition within a given research area or line of investigation.

The fact that components are contained within mechanisms ensures that components are of a smaller size than the mechanism itself, but, as the

examples in the last paragraph reveal, not all entities at a given level will be of the same order of magnitude. The level at which a particular entity resides depends on the role it is playing in the mechanism. A heart is a critical part of the circulatory system of a surgeon, which in turn plays a critical role in the ability of the surgeon to perform surgery. But when the surgeon is performing heart surgery, she is interacting with the patient's heart – holding it, reconnecting it, etc. In that interaction the patient's heart is at the same level as the surgeon performing the operation, whereas the surgeon's own heart is at a lower level. Likewise, in oxidative phosphorylation, there are protons in the atoms that comprise the lipid molecules of the inner mitochondrial membrane as well as protons being pumped across the membrane. The protons pumped across the membrane interact with it (as illustrated later in Figure 6.9), but protons in the lipid molecules are at a lower level than the protons being pumped.

So far I have emphasized that in a mechanistic account of levels it is the components of a mechanism that are the denizens of a lower level. In contrast, it is the mechanism as a whole and the things with which it interacts that inhabit the higher level. These interactions with other things may greatly influence the behavior of the mechanism and many of these things may themselves be fruitfully construed as mechanisms. In some cases a mechanism may be part of an organized system in which its behavior is coordinated with that of other entities to perform yet another function. In that case, the first mechanism is a component of the yet higher-level mechanism. If the organization in the higher-level mechanism results in the imposition of constraints on the behavior of the first mechanism, such embedding of a mechanism into a higher-level mechanism can be highly relevant for the scientist trying to explain the operation of the initial mechanism. Moreover, just as one can go up from a given mechanism to a higher-level mechanism, one can proceed down from a component of a mechanism to yet lower-level components. A mechanistic account thus gives rise to a cascade of levels.

Given the differences in the way levels are characterized, it is not surprising that the mechanistic account of reduction looks very different from traditional philosophical accounts of reduction which hold that higher-level theories are reduced by logically deriving them from lower-level theories (Nagel, 1961; Causey, 1977). For such a derivation to succeed, all the information required to generate the higher-level must be contained in the lower-level theory. If this is the case, a successful theory reduction renders the derived theory superfluous – everything is explained equally well with the lower-level theory. With mechanistic explanations, accounts of the lower level do not offer a complete theory. None of the components, alone, generates the phenomenon.



Even the whole collection of components does not produce the phenomenon except when appropriately organized. But it is precisely by organizing the parts and their operations that a higher level is constituted, thus providing a bridge linking the level of the components to the level of the mechanism.

It is important to emphasize that a mechanistic account, while appealing to parts and their operations, also appeals to the functioning of the mechanism as a whole and how, in virtue of its function, it interacts with entities in its environment. There is, therefore, a critical difference in what is investigated at the lower and higher levels. At the higher level investigators characterize the functioning of the mechanism as it is situated in its environment. Pasteur, for example, characterized yeast cells switching from aerobic to anaerobic metabolism depending on the presence of oxygen in their environment. Biochemical studies, which were not available to Pasteur, could open up the mechanism and reveal the components and their operations that made yeast capable of this switch. The biochemical processes, however, only suffice to explain how the function is performed *in the particular context*. Even after the component parts and operations are identified, it remains the case that the mechanism as a whole in its context performs a function: Switching from aerobic to anaerobic metabolism remains something done by yeast cells (or at least the whole glycolytic system in them). The upper level is far from superfluous – explanation requires both the account of what is happening at the higher level and the account of the components of the mechanism. In this respect, the kind of reduction that arises with mechanistic explanations is also compatible with recognizing the autonomy of higher levels of organization. The operations at a level are unique to the level and must be investigated with the appropriate tools.

The role of organization in generating higher levels is made clear by considering the point of view of an engineer. When an engineer faces a task, what she must do is draw upon the operation of components she already has available and organize them in a new way to accomplish the task. (She may have to iterate this strategy by decomposing what she is trying to accomplish into yet finer-grained operations, some of which can be performed by existing components, but others requiring another level of decomposition so as to design components that can perform those subsidiary tasks.) When she has finished, she has built something new, perhaps something for which she could secure a patent. We would not expect the patent office to deny her a patent because all of the components were already known to her – they were also known to the others who failed to have the insight needed to develop the new mechanism. What she has done is to build a mechanism that performs a new function.

This appeal to engineering highlights the fact that mechanisms perform different functions than their components and emphasizes the importance of organization of parts and operations in accomplishing the new function. If the engineer relied on off-the-shelf components in building her mechanism, then the only thing she added was organization. It is the only thing she added, but it is far from trivial. It is for discovering the way to organize parts and operations that the engineer earns her patent. But there is a further factor that successful engineers take into account – the context in which the mechanism performs its function. One common way in which engineered products fail is when they are put into operation in contexts other than those for which they were designed. The wrong context can hinder a mechanism's operation whereas the appropriate context may provide things which are co-opted in the mechanism's operation. This is even more true of biologically evolved mechanisms than humanly engineered ones. Evolution is an opportunist, and if something can be relied upon in the mechanism's environment, then it doesn't have to be generated by the mechanism. Vitamins provide just one well-known example. Because our ancestors could generally count on the availability of vitamins in their foods, there was no evolutionary pressure for us to retain the ability to synthesize them. Insofar, however, as such environmental factors are necessary for the functioning of the mechanism, mechanistic explanations need to focus on the mechanism's context, not just its internal configuration.

## 6. ORGANIZATION: FROM CARTESIAN TO BIOLOGICAL MECHANISMS

I described previously how, in reviving the mechanical philosophy, Salmon significantly expanded the toolbox of features in terms of which mechanisms could be understood. Instead of just shape and motion (Descartes' features), modern mechanists can appeal to such things as gravitational and electromagnetic fields. In biology the comparable expansion brings in chemical bonds and catalysis. But biological mechanisms require expanding the toolbox Descartes supplied in yet a different way, one that focuses on organization of parts and operations. The appeal to engineering showed how organization enables a mechanism to perform functions which its parts alone cannot, but the organization of biological systems is distinctively different from the organization typically exhibited in humanly engineered systems. When humans think of putting components together we usually think of linking them together in series so that each component operation sends its product to the next component operation in the series. From biological systems, however, we have

learned of more complex modes of organization that achieve rather surprising results.

The significance of organization for biological mechanisms was brought home in the nineteenth century by challenges from biologists who denied that mechanisms could account for the phenomena of life. These biologists, known as *vitalists*, highlighted ways in which biological systems function differently than non-biological systems. Xavier Bichat (1805) is an important example. In many respects, Bichat was pursuing a program of mechanistic explanation. He attempted to explicate the behavior of different organs of the body in terms of the tissues out of which they were constructed. He decomposed these organs into different types of tissues that varied in their operations and appealed to the operations of different tissue types to explain what different organs did. But when Bichat reached the level of tissues, he abandoned the mechanist program. This was because tissues exhibited two features which he thought defied mechanistic explanation. First, tissues are indeterministic in their response to external stimuli. In contrast, machines as he conceived them always respond the same when presented with the same stimulus. Second, they seem to resist those environmental forces that threaten them. It is relatively easy to take apart or interrupt a machine and stop its operation, but living tissues are often difficult to thwart or kill. These differences, Bichat thought, undermined any hope of providing a fully mechanist account of living tissues.

Several decades later Claude Bernard (1865) sketched a mechanistic answer to Bichat. It involved identifying principles of organization in living systems that could account for the features to which Bichat pointed. Bernard presented his answer by distinguishing two environments. The term *environment* is typically used only for the first – the external environment in which the organism as a whole lives. Bernard proposed that biologists also need to consider the local environment of each organ (mechanism) within a living organism. This he termed the *internal environment*. Component mechanisms in the living organism interact directly with this internal environment, not with the external environment. This internal environment provides a buffer between the conditions in the external environment and the local mechanisms. The various organs respond to the conditions of the internal environment, and these responses might be quite deterministic – when conditions in the internal environment differ, the organs predictably behaved differently. For example, decreased glucose levels in the blood could lead to lowered metabolic activity in somatic tissues or reliance on a different metabolite.

How did the internal environment serve as a buffer? Bernard proposed that each mechanism within the organism monitored an aspect of the internal environment and operated to maintain that feature of the internal environment

in a constant condition.<sup>15</sup> The joint consequence of different internal mechanisms each operating to maintain a given feature of the internal environment is that each mechanism enjoys a stable environment in which to operate. Hence, each is buffered from conditions outside the organism – as external conditions begin to change conditions inside the organism, the appropriate mechanism registers the change in the internal environment and acts to restore it to its normal condition. Because of this internal buffering, the various other mechanisms, and the organism as a whole, do not show regular causal responses to perturbations in the external environment and so are indeterministic in the manner Bichat noted. Moreover, insofar as each component mechanism is successful in performing its activity as needed to keep the internal environment constant, the organism continues to maintain itself in the face of conditions that would seem to have the potential for destroying it. Thus, he can explain why organisms seem to resist external factors that would seem deadly to them.

Bernard's account leaves some important features of biological mechanisms unexplained. Perhaps the most important feature left unexplained is *how* it is that component mechanisms maintain "the constancy of the internal environment" (Bernard, 1878a, p. 113). In order to do this, though, each component mechanism must itself have a complex structure, including components that serve to detect when the operation is needed and a means to activate the mechanism so that the operation is performed when and only when it is needed. Walter Cannon (1929) introduced the term *homeostasis* (from the Greek words for *same* and *state*) for the capacity of living systems to maintain a relatively constant internal environment. He also sketched a taxonomy of strategies through which animals are capable of maintaining homeostasis. The simplest involve storing surplus supplies in time of plenty, either by simple accumulation in selected tissues (e.g., water in muscle or skin), or by conversion to a different form (e.g., glucose into glycogen) from which reconversion in time of need is possible. Cannon noted that in most cases such conversions are under neural control. A second kind of homeostasis involves altering the rate of continuous processes (e.g., changing the rate of blood flow by modifying the size of peripheral blood vessels to maintain uniform temperature).

Cannon's particular interest was the role of the autonomic nervous system in regulating supplies or processes, but during the same period biochemists

<sup>15</sup> Bernard, for example, says, "all the vital mechanisms, however varied they may be, have only one object, that of preserving constant the conditions of life in the internal environment" (1878a, p. 121, translated in Cannon, 1929, p. 400).

were discovering the ubiquity of cyclic organization of biochemical processes and the capacity of such cycles to provide auto-regulation of these processes. One of the first cycles proposed was by Otto Meyerhof, drawing upon the research of Archibald Hill on two periods of heat generation in muscle contraction, one accompanying formation of lactic acid during muscle contraction itself and the second accompanying the subsequent disappearance of lactic acid during a recovery phase (Hill, 1910; Hill, 1913). Meyerhof (1920; 1924) proposed a cycle, which he termed the *lactic acid cycle*, in which approximately three-quarters of the lactic acid formed during the anaerobic contraction phase was resynthesized to glycogen at the expense of the remaining molecules of lactic acid, which were further oxidized to carbon dioxide and water. As I will discuss in the next chapter, other proposals for cycles soon followed and as their ubiquity became clear, the question of why they occur so frequently arose. One reason is that they provide a means of regulation via such procedures as feedback control.<sup>16</sup> We saw an example of this in Figure 2.3. As another example, the cycle in which ATP is broken down to ADP and  $P_i$  in the course of muscle work and resynthesized during energy metabolism provides a means by which energy metabolism can be regulated to proceed only when ATP is needed for work.

Systems organized to regulate themselves are not unique to biology. Maintaining constant conditions is also important in many human artifacts and engineers have invented ways for doing so. Water clocks, for example, required that the water-supply tank be maintained at a constant level so that water ran out at a constant rate, and a feedback control system for such clocks was developed by Ktesibios in approximately 270 BCE. Windmills need to be pointed into the wind, and British blacksmith E. Lee developed the fantail as a feedback system to keep the windmill properly oriented. A temperature regulator for furnaces was developed by Cornelis Drebbel around 1624. Although each of these used a version of negative feedback, they were isolated developments limited to the machines in which they were utilized.<sup>17</sup> The governor developed by James Watt for his steam engine attracted much

<sup>16</sup> Mercer described the general consequence of such organization, which he attributed to James Danielli's "generalized cell theory": "The whole complex of enzymatically controlled reactions – in which the product of one reaction forms the substrate for the next, and often a final product is returned (or fed-back) to re-enter a cycle of reactions at another point – constitutes a system in dynamic equilibrium buffered against change, so long as material and energy is fed into it" (1962, p. 50).

<sup>17</sup> Harold Black's development in 1927 of negative feedback as a means of controlling feedback distortion in amplifiers such as those used in telephones – by feeding back the signal from the amplifier so as to compare it with the input signal – illustrates how the principle had to be rediscovered in each individual case.

more attention, in large part because it became the focus of James Clerk Maxwell's (1868) mathematical analysis of control systems using differential equations. Watt's challenge was to maintain the steam engine at a constant speed despite the fact that the loads upon it varied (as, for instance, different sewing machines attached to it in a textile factory would come on and off line). His solution was to attach a spindle to the drive shaft and then attach moveable arms to the spindle (see Figure 2.4). Centrifugal force would cause these arms to open up more the faster the engine was running. To these angle arms Watt attached a linkage mechanism that would reduce the opening of the steam valve the more the arms opened and increase the opening the more they dropped. Whenever the engine ran too fast, the arms would rise and that would cause the valve to close, slowing the engine down. Whenever it ran too slowly, the arms would drop, and that would cause the valve to open, speeding the engine up.

As technological systems developed for which control was critical, negative feedback came to be recognized as a powerful tool. It provided the foundational idea for the cybernetics movement,<sup>18</sup> of which Norbert Wiener was a driving force. Wiener had interest and training in biology before earning his Ph.D. in mathematical logic and making important contributions to pure mathematics. Because he often collaborated with biologists, his inspiration for emphasizing cyclic organization probably lay as much in biology as in mathematics and engineering. During World War II he and Julian Bigelow took on the challenge of developing a control system for antiaircraft fire (the challenge stemmed from the fact that the airplane moves a considerable distance in the approximately one-minute interval that it takes a projectile to reach it). They recognized that the problem was not essentially different from the problem an animal faces in moving its limbs and were inspired to employ negative feedback. Thus, their strategy was to predict from radar information the future location of the plane but then correct the prediction based on the difference between the actual and predicted location at the next timestep. They found, however, that if the feedback signal was at all noisy and the system responded too quickly, feedback caused it to go into uncontrollable oscillations. Wiener and Bigelow consulted with Mexican physiologist Arturo Rosenblueth, who reported similar behavior in human patients with damage to the cerebellum. In both cases, they concluded, averaging techniques had to be invoked to dampen the response.

<sup>18</sup> Indeed, Wiener was led to the term cybernetics while thinking about Watt's governor. Looking into the etymology of the word *governor*, Wiener followed the path to the Greek word for governor, *kybernan*, and encountered the Greek word *kybernetes*, steersman.

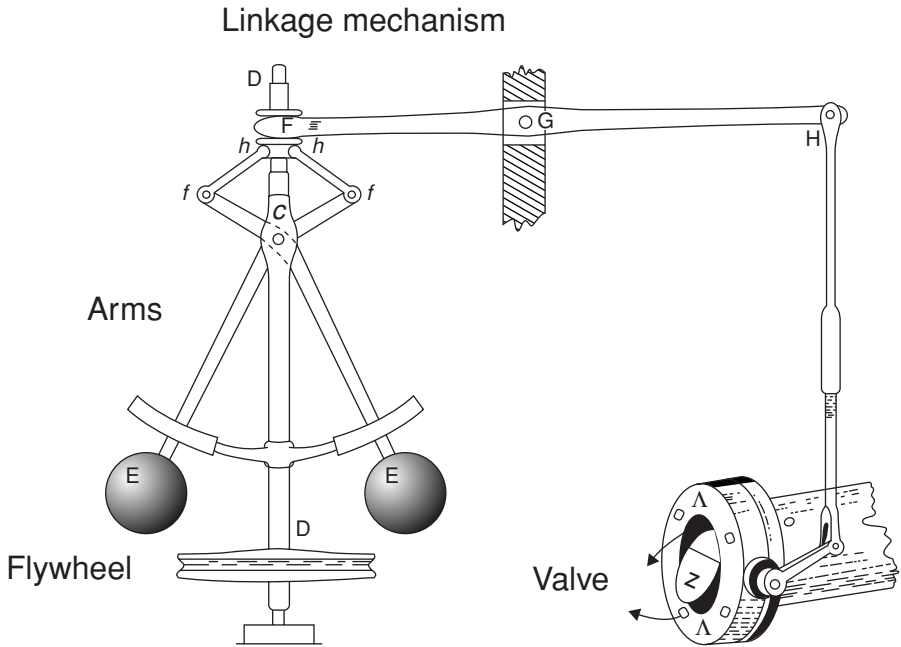


Figure 2.4. The governor James Watt designed for his steam engine. An upright spindle is attached to the flywheel. Connected to it are arms with balls which, by centrifugal force, will move further out the faster the flywheel turns. There is a linkage mechanism which ensures that the valve closes as the balls move further out, and opens as they move closer in. This ensures that steam will flow at the rate needed to keep the flywheel moving at the desired speed. Drawing reproduced from J. Farley (1827), *A treatise on the steam engine: Historical, practical, and descriptive*. London: Longman, Rees, Orme, Brown, and Green, p. 436.

Wiener and his colleagues did not view their contribution only as an advance in technological design (in fact, Wiener's design failed due to the limitations of available hardware in which to implement it, and his contract with the Department of Defense was terminated). Rather, they thought they had articulated a basic principle of purposive and teleological behavior, whether in animals or machines. Accordingly they published their results in the journal *Philosophy of Science*.<sup>19</sup> They acknowledged that the concept of teleology

<sup>19</sup> On their account, *purposeful* is more general than *teleological*. "The term purposeful is meant to denote that the act or behavior may be interpreted as directed to the attainment of a goal – i.e., to a final condition in which the behaving object reaches a definite correlation in time or in space with respect to another object or event" (Rosenbluth, Wiener, & Bigelow, 1943, p. 18). They then invoked feedback to differentiate teleological from non-teleological behavior: "Purposeful active behavior may be subdivided into two classes: 'feed-back' (or 'teleological') and 'non-feed-back' (or 'non-teleological')" (p. 19).



was largely discredited because it suggested final causation or future outcomes directing earlier events, but maintained that the kind of purposeful behavior guided by feedback was sufficiently important to resuscitate the term. Wiener and his collaborators were so impressed with the potency of negative feedback that, with support from the Macy Foundation, they established twice yearly conferences. Initially the conference was known as the Conference for Circular Causal and Feedback Mechanisms in Biological and Social Systems, but after Wiener (1948) coined the term *cybernetics*, the name was changed to the Conference on Cybernetics.

Teleology is an important feature of biological systems, and cyclic organization is important for designing systems that can achieve ends on their own – whether the system is biological or an artifact designed by humans. There is, however, another feature of biological systems that is critical, but that has not received much attention to date by philosophers focusing on mechanism. The ability of mechanisms to function, and especially to regulate themselves, depends critically on the particular ways in which their parts are organized. Being organized, though, is not the natural state of matter. Rather, disorder, or equal distribution, is the state to which physical matter tends. This is the import of the second law of thermodynamics – that in a closed system, entropy (disorder) increases. The only truly stable state, referred to in thermodynamics as the state of maximum entropy, is one in which the components are equally and randomly distributed. When this obtains, the system is at equilibrium.

Given the second law of thermodynamics, biological mechanisms pose a puzzle. As highly organized systems, they are far from thermodynamic equilibrium. According to the second law, this organization should break down and such mechanisms should approach equilibrium. How is it possible to keep a system far from thermodynamic equilibrium – that is, to keep it organized? Part of the answer is that biological systems and their environment are not closed systems; energy is always entering from the outside. However, this energy must be appropriately directed to maintain organization. With human-made mechanisms, the most common means of maintaining organization is to rely on an external repair system, often a human being. Much like the original builder of a mechanism, the repair person utilizes energy, originating outside the system, to reorganize the components when the order between them breaks down and to replace components when they internally break down (become disorganized). In general, however, biological systems do not have external repair people to come in and expend energy to rebuild them. They must do it themselves. How is this possible?

A variety of twentieth-century theorists made important theoretical contributions that provide in basic outline an account of how biological systems



maintain themselves far from thermodynamic equilibrium. One contribution is the recognition of self-organizing chemical systems such as the Belousov–Zhabotinskii (B–Z) reaction in which simple component reactions together give rise to complex patterns. Ilya Prigogine coined the term *dissipative structures* for constantly changing, highly organized systems (Nicolis & Prigogine, 1977). Such order depends on a constant flow of energy through the system, which is employed to create order but then dissipates as heat. Dissipative structures such as the B–Z reaction, however, can neither procure for themselves the needed energy flow to maintain themselves far from equilibrium nor the chemicals needed to continue the reactions. Hence, they soon terminate. In the early 1970s, Francisco Varela and Humberto Maturana introduced the term *autopoietic machine* for systems such as living organisms that maintain themselves by producing their own components. They emphasized that the components of such systems “1) through their interactions and transformations continuously regenerate and realize the network of processes (relations) that produced them; and 2) constitute it (the machine) as a concrete unity in the space in which they exist by specifying the topological domain of its realization as such a network” (Varela, 1979, p. 13).

Both characteristics of autopoietic machines deserve comment. First, to regenerate themselves, such machines require machinery that is directed toward making their own components. Such machinery requires a source of energy and raw ingredients for constructing itself. Hence, the system must be an open system situated between an energy source (a more structured, less equally distributed, state of matter) and an energy sink (which is less structured). But crucially it must capture this energy in a useable format (e.g., gradients, chemical bonds) so that it is available for energy demanding operations. Second, the smooth operation of the system depends upon the organization of the components so that different operations are orchestrated (e.g., energy and raw materials are delivered to the location where new parts are synthesized). Minimally, this requires some way to keep needed constituents together and separated from the environment in which the system is situated. That is, the system requires a partially porous boundary (e.g., a semi-permeable membrane) that allows selected materials to cross. Moreover, the system itself must be able to control what is admitted and what is expelled from the system. Ultimately, because the concern is with a system that remakes itself, this boundary needs to be one of the system’s own making.<sup>20</sup>

<sup>20</sup> For a stimulating and provocative development of how a metabolic and a membrane system, together with a control system, can result in a chemical system capable of exhibiting the basic features of living systems, see Gánti (2003).

Varela uses the term *autonomy* for systems that define and maintain their own boundaries. Ruiz-Mirazo, Peretó, and Moreno provide a useful characterization of an autonomous system as

a far-from-equilibrium system that constitutes and maintains itself establishing an organizational identity of its own, a functionally integrated (homeostatic and active) unit based on a set of endergonic-exergonic couplings between internal self-constructing processes, as well as with other processes of interaction with its environment. (2004, p. 330)

Autonomy in this sense is a feature of any living system. Although each mechanism does not have to be autonomous, it must be part of a system that is.

The requirement of being built up through self-organizing processes also provides a means to address a key feature of Bichat's critique of mechanism with which I began this section. Bichat contended that living systems resist death. In fact, this generally takes the form of adaptive change over time so as to deal with novel environments. Understanding the adaptive capacities of biological organisms is challenging. Developmental theorizing, in both biology and psychology, has been driven by the polar positions of nature and nurture. Advocates of a nurture position maintain that the organization was not substantially prespecified, but rather resulted from interactions with the environment. But advocates of the nurture position have faced apparently insurmountable obstacles. The end-product of both biological and psychological development is a highly structured state. Ensuring the achievement of such structure seems extremely problematic if everything must be directed from the environment. The nature position seems improbable in its own right, requiring a vast amount and efficacy of innate information.

Within the sphere of cognitive development, the psychologist Jean Piaget proposed an alternative to the polarizing options of nature and nurture – a position he referred to as *constructivism*. He introduced this alternative as follows: “The essential functions of the mind consist in understanding and in inventing, in other words, in *building up structures* by structuring reality” (Piaget, 1971, p. 27). Piaget's attempt to find a middle path between nature and nurture, however, turned out to be bitterly contested. The philosopher of psychology Jerry Fodor objected, “It is *never* possible to learn a richer logic on the basis of a weaker logic, if what you mean by learning is hypothesis formation and confirmation” (1980, p. 148). Fodor's idea was that before one can test a hypothesis one must be able to formulate it, and that entails the capacity to represent it. Thus, the capacity to represent any hypotheses that one will ever test must be innate, making the native endowment very

powerful. Fodor's claim is premised on the condition that learning consists of hypothesis formation and confirmation, which is itself contentious. To contest it, however, one must offer an alternative, and this has proven difficult. The notion that autonomous systems must be built ultimately out of self-organizing systems, however, perhaps offers the foundation for building an alternative to Fodor's assumption. Self-organizing systems are able to construct richer structures than they start with (see Elman et al., 1996, for a suggestive account of cognitive development).

This points us to yet another contrast with the notion of mechanism adequate for nonliving systems. The engineering strategy, as we have seen, is to build more complex systems out of simpler components by imposing organization upon existing components. This is a very productive strategy, but it depends upon the capacity of the engineer to propose new forms of organization. Biological systems, however, cannot rely on external engineers. (Given natural selection, they can of course rely on selective retention of chance variation, but the generation of new useful modes of organization through chance variation of already evolved systems is extremely uncommon.) Current modes of organization in artifacts are usually located near local peaks on adaptive landscapes so that small variations are likely to be detrimental. Fortunately, however, biological systems can utilize a different approach. If biological systems are self-organizing, they are also capable of self-reorganization. This means that they can do more than impose a new organization on existing components. Insofar as these self-organizing capacities remain active and not frozen within the system, the system can also generate from within new components capable of performing new activities.

At present we have only limited models of such self-organizing of new components with new capacities within living systems. These include cases of plasticity in nervous systems. Nervous systems wire themselves and components emerge and take on specific functions partly as a result of the neurons that synapse on them. As a result, tissue which in a normally developing individual will serve visual processing tasks will, in a brain in which there are no inputs from eyes, take on the processing of auditory or other sensory inputs (Pascual-Leone & Hamilton, 2001). And areas in the somatosensory cortex that process information from fingers will reorganize if the input changes if, for example, an investigator binds together two digits so that they can no longer respond differentially (Merzenich et al., 1990). Although the details are less clear, such processes provide the most plausible explanation for the existence of an area in the brains of literate humans that has apparently become specialized for visual letter patterns (Petersen et al., 1990). Such a processing area could not have been specified in a genome when that genome was

selected in organisms that did not read written texts. It must be the product of self-reorganization of prior processing areas in organisms that had to adapt to and become adept at reading written characters.

## 7. DISCOVERING AND TESTING MODELS OF MECHANISMS

So far I have focused on articulating what are mechanisms and mechanistic explanations so as to make it clear what biologists, including cell biologists, are pursuing in their investigations. I have not said anything about the processes of discovering and evaluating claims about mechanisms. Because these are the prime activities of scientists, I turn next to the question of what can be said about these processes.<sup>21</sup>

The very conception of a mechanism sets the goals for its discovery – the investigator must identify the parts of the mechanism, determine what operations they perform, and figure out how they are organized to generate the phenomenon. Earlier I introduced the notion of *decomposition*, but emphasized the conceptual side of decomposition. To develop a model of a mechanism, a theorist decomposes it conceptually into parts and operations. But discovering a mechanism usually results from experimental engagement<sup>22</sup> with the mechanism. I differentiated two ways researchers decompose mechanisms – structurally or functionally – depending on whether they focus on component parts or component operations. As we turn to experiments, these two types of decomposition now correspond to two types of experimental

<sup>21</sup> Logical empiricists typically rejected the possibility of philosophical analysis contributing to understanding what Reichenbach (1966) called the context of discovery and instead focused on the context of justification, where it was thought that logic could characterize the relation of evidence to hypotheses. Interest in discovery was rekindled around 1980 (Nickles, 1980a; Nickles, 1980b). See in particular Darden (1991), who focuses on discovery in the context of theory revision. When mechanisms are the focus, it turns out that quite a bit can be said about scientific discovery.

<sup>22</sup> Another set of experiments aim simply to establish the phenomenon for which a mechanistic explanation is sought. Experimentation designed to assess the phenomenon itself generally does not try to take the mechanism apart, but rather seeks to determine regularities in its behavior by establishing relations between inputs to the mechanism, conditions of its operation, and output. This involves measuring the values of relevant variables and, often, the manipulation of some variables. Lavoisier and Laplace (1780), for example, established the similarity of respiration by animals and ordinary combustion by placing respiring organisms or burning coal in a calorimeter and measuring the amount of carbon dioxide produced and of ice melted. This only involved setting up appropriate conditions for measuring the relevant variables. Pasteur (1861), on the other hand, determined that fermentation was an anaerobic phenomenon by manipulating the presence or absence of oxygen and showing that fermentation was suppressed in the presence of oxygen.

procedures – those directed at the identification of component parts and those focused on component operations.

Although the goal of discovering the component parts and operations of mechanisms is clear, actually succeeding often requires great ingenuity. Mechanisms generally do not directly reveal how they operate. All we observe is their functioning and often those observations require careful interpretation. Even when the internal operations involve changes in an identifiable substrate, the intermediate states of the substrate can be hidden because in a well-organized machine individual operations flow smoothly into each other without leaving any trace of their intermediate products. Think again of engineering, where the ultimate goal is to design a device of which the user is not even aware and which operates when needed without any intervention by the user. In the early days of a new technology this desiderata is often not realized. Drivers of early automobiles had to understand how they operated so as to be able to repair them after their frequent failures. Users of early computers had to understand their operation at the machine level, or at least at the level of the operating system, in order to get their programs to perform as desired. As computers matured such knowledge became unnecessary and the details of the operation were increasingly hidden from view (e.g., behind a smoothly functioning windowing system and higher-level programming languages).<sup>23</sup> An engineer who wants to emulate a competitor's system but lacks access to the design specifications often faces a serious challenge of reverse engineering – taking the system apart and trying to figure out how it was put together. This is the situation confronting investigators of biological systems who must rely on experimental manipulations to take apart and reveal the parts and operations within them.

### *Identifying Working Parts*

Some experimental inquiries are designed simply to separate the parts of a mechanism. It is important to emphasize that the parts into which researchers seek to decompose a mechanism are ones which perform the operations that figure in the functional decomposition. The majority of ways of structurally decomposing a system will not result in parts that perform operations. As Craver (forthcoming) notes, one might dice any system into cubes, but these cubes do not individually perform operations in terms of which one can explain the phenomenon. To reflect this fact, I will, following Craver, refer to

<sup>23</sup> Herbert Simon (1996) made this point in the context of bridge design. When a bridge is functioning as intended, it reveals little of its design principles.

the relevant component parts as *working parts*. Although the goal is to find working parts, it is possible to decompose a system structurally independently of actually being able to determine the operations the various components perform. This involves, for example, appraising that component structures are likely to be distinct working parts on other grounds.

Let's begin with interventions that are designed to identify the parts of a mechanism. One way is to apply enough force that it breaks down into separate components that can then be further investigated as potential working parts. Deciding what counts as "enough" force is a challenge. Some ways of applying force will completely obliterate the components of interest – those that perform the operations within the mechanism. One could, for example, cut a brain into tiny cubes or homogenize a cell in a Waring blender. This is not problematic if the units of interest are at a sufficiently small scale (individual neurons or ganglia in the brain or individual enzymes in the cell) to survive intact. If the task of interest relies on larger units (brain regions or cell organelles), the intervention will have destroyed the working parts. There is a further challenge, though. Assuming that a researcher has arrived at a method that delivers the right amount of force to get parts of about the right size, how would she know whether these are the parts that divide the system at its *joints*? Consider performing an anatomical dissection. How does the anatomist know whether she has dissected out a working part (cut at the joints) or has cut within a part? Sometimes components (e.g., organs such as the heart, liver, and pancreas) exhibit clear boundaries and integrity. This does not entail that they are functionally relevant units, but it provides a good clue.

Physical separation is not the only possible strategy. In some fields researchers have a means of viewing internal components. Cytologists, for example, were able to distinguish cells as parts of organs, and nuclei as parts of cells, using the light microscope and smaller parts using the electron microscope. In Chapter 4 I will discuss the epistemic challenge in determining whether such visualizations are reliable. For now what is important to note is that visualization provided a partitioning based on appearance that was useful for many purposes, but needed to be corroborated by other methods to confirm correspondence to working parts. The history of attempts to identify working parts in the brain illustrates the challenge (Mundale, 1998). In highly convoluted brains such as the human brain, the existence of sulci and gyri seemed to provide natural boundaries between regions of cortex. Although these are still used as reference points in locating regions in the brain, it is now recognized that the process of folding that gives rise to sulci and gyri does not respect function as much as expected. Hence, other evidence is needed to identify the working parts. Brodmann (1909/1994) as well as several other

investigators at the beginning of the twentieth century invoked other criteria such as the type of neuron found in an area and the thickness of different layers of cortex to demarcate regions. Although Brodmann lacked any means of linking neural operations with these areas, he clearly hoped that these criteria would differentiate working parts. Subsequently a variety of other indirect criteria, such as topographical mapping, have been used to refine such maps (van Essen & Gallant, 1994). The justification for using such indirect criteria is the assumption that they track features (e.g., morphologically distinct types of neurons) that should matter operationally.

Ultimately, it is the integration of structural decomposition with functional decomposition that provides real answers to the question of which parts are working parts. In subsequent chapters we will encounter a variety of methods employed for structural decomposition of cells and also the advantage attained when cell biologists integrated these results with those from functional decomposition.

### *Identifying Component Operations*

Turning now to functional decomposition, here the strategy is to start with the overall function of the mechanism and figure out what lower-level operations contribute to achieving it. These operations are characterized differently in different sorts of mechanisms. In biochemistry, the typical operation is a chemical reaction catalyzed by an enzyme (active part) that transforms a substrate into a product (the passive parts). The biochemical system that performs glycolysis in cells, for example, catabolizes glucose to pyruvate via a series of such reactions, each of which may oxidize or reduce a substrate, add or remove phosphate groups, etc. Often it is possible to determine, at least to a first approximation, what the internal operations of a mechanism are without knowing what active parts perform these operations (the passive parts do need to be known, though, because the identity of the operation depends upon what is changed). In the case of glycolysis, biochemists were able to determine the chemical reactions that realized this function and to designate a responsible agent (enzyme) for each reaction without knowing the chemical structure of the agents (Bechtel, 1986b). Instead of direct access to the structures operative within the system, such decompositions rely on cleverly designed perturbations of the functioning of the system that provide clues to the component operations. In such cases, then, functional decomposition precedes structural decomposition.

Some experiments designed to identify internal operations involve manipulation and measurement of variables without explicitly going inside the



mechanism. In such cases each manipulation should be precisely designed to target a specific internal operation, such that a distinctive effect on the behavior of the whole system will be obtained.<sup>24</sup> In most cases, though, experiments designed to determine the internal operations in a mechanism involve intervening or measuring changes within the mechanism. The strongest evidence may arise when an investigator can directly intervene on and measure effects in the operation of a component in the mechanism. Sometimes this can be approximated by separating a component from a mechanism, experimenting on it under realistic conditions to determine what it could do, and fitting that information into a coherent account of how the whole mechanism performed its function. Biochemists, for example, may attempt to isolate from cells what they take to be the relevant enzymes and cofactors for a given reaction and then conduct experiments on the isolated components. Frequently it is not possible to isolate the factors responsible for an operation within a mechanism in this manner, in which case researchers have to settle for more indirect routes for gaining information.

Regardless of whether it is structural or functional components that provide the initial target, two strategies for determining whether and how they figure in a mechanism are excitation and inhibition. The effects of these manipulations may then be registered in the overall behavior of the mechanism or more locally within the system. By inhibiting or removing a component, investigators force the system to operate without it. The hope is that from the altered behavior of the mechanism they can figure out what part or operation was involved. This is not always easy, as Richard Gregory argued:

Although the effects of a particular type of ablation may be specific and repeatable, it does not follow that the causal connection is simple, or even that the region affected would, if we knew more, be regarded as functionally important for the output – such as memory or speech – which is observed to be upset. It could be the case that some important part of the mechanism subserving the behavior is upset by the damage although it is at most indirectly related, and it is just this which makes the discovery of a fault in a complex mechanism so difficult. (1961, p. 323)

Moreover, in many cases, the removal of a component part produces a whole cascade of effects, leading the mechanism to no longer carry out anything like

<sup>24</sup> Reaction time studies used in cognitive psychology provide an illustrative example. Different proposals for internal operations were linked with different timing expectations so that actual reaction time measures could decide between them.



the function normally associated with it. Removing a transistor from a radio, to use another of Gregory's examples (Gregory, 1968), may cause the radio simply to hum, perhaps giving the misleading suggestion that the transistor was a hum suppressor. Nonetheless, inhibiting or removing a part sometimes provides important clues as to what operation it performed. For example, if removing a particular enzyme inhibits a reaction and thereby an entire function (e.g., fermentation) and also leads to the build-up within the system of another substance, then it provides evidence both that the reaction contributes to that function and that the accumulating substance was an actual intermediate (or the product of a further reaction on that intermediate). Although single clues of this type are subject to multiple interpretations, the results of multiple different inhibitions within the system often provide insight into what operations different parts are performing.

The complement to inhibiting an operation is to stimulate it. For example, investigators can supply additional amounts of what they think is an intermediate substance in a chemical pathway and determine whether that results in a greater output from the pathway. This too can lead to anomalous results, either because the substance did not reach the right point in the pathway or because, without other components cooperating appropriately (perhaps via feedback loops), the operation was blocked. Often the stimulation technique can be fruitfully combined with an inhibition. For example, investigators might inhibit a reaction by removing the enzyme needed to obtain a particular product but also supply the product from outside. If the ultimate product of the reaction pathway continues to be produced normally, then the investigators can have high confidence that the enzyme they removed was part of the normal pathway.

Both inhibition and stimulation studies involve manipulating a component in the normal operation of the mechanism and detecting the consequences. Sometimes it is possible to observe the operation of the components directly. One example is single neuron recording in neuroscience, in which an electrode is inserted near a neuron and then various stimuli are presented to see which types produce increased (or decreased) spiking. Another example is radioactive tracer experiments, in which researchers label a substrate with a radioactive component and then detect where the radioactivity later appears within the mechanism. When such recordings are possible, they often provide the most compelling evidence of the sequence of internal processes. They do not, however, reveal the nature of the operation involved at each locus; hence, recording studies usually must still be combined with inhibition or stimulation studies to determine the operations.

*Localizing Operations in Parts*

I noted earlier that sometimes researchers are able to make greater progress in identifying parts than in specifying operations, or vice versa. One reason this happens in the history of science is that the techniques for carrying out one form of decomposition are developed in a different field and at a different time than those for carrying out the other form of decomposition. We will see in subsequent chapters that this was often the case with regard to mechanisms in the cell. Biochemists, focusing on functional decomposition into component operations, identified biochemical reactions. Cytologists, focusing on structural decomposition into component parts, identified those organelles that were within the resolving power of their microscopes. It was engagement between these fields that gave rise to cell biology. Increasingly, parts and operations in what previously had been a terra incognita were identified and aligned to provide new mechanistic accounts (e.g., it was found that one part of the mitochondrion – its inner membrane – performed the operation of maintaining a key energy gradient). Equally important, it was learned which of the newly discovered mid-level structures housed which lower-level biochemical reactions. For example, the cytosol was found to be the site of the reactions involved in glycolysis. This kind of alignment, crossing levels rather than remaining at a single level, at the very least provided a more complete account of the cell. Often there were further benefits, as important hints and constraints suggested by interlevel alignments led to improved accounts at one or more of the levels involved. At various points in the chapters that follow, I use the term *localization* to refer not only to alignments within a single mechanism but also to those that crossed levels in the cascade of mechanisms being uncovered by cell biologists.

The ability to link operations with parts often depends upon developing experimental procedures that can provide appropriate linking information. In the case of cell biology, the ability of electron microscopy to identify the structures found both in whole cells and in preparations of working parts isolated by cell fractionation provided the primary evidence relating biochemical operations with the organelles of the cell. As discussed in Chapter 4, an important contribution of such connections is that they provide a means of corroborating each decomposition. Linking a component operation with an independently identified component part provides evidence that both really figure in the mechanism. Failure to link operations with parts, on the other hand, can be grounds for doubting the existence of either the part or the operation. As we shall see, cell organelles were often claimed to be artifacts until they were shown to be the locus of important operations in the cell. Linking an

operation to a structure also provides evidence that the hypothesized operation actually occurs.

Although I have stressed the possibility of either decomposition proceeding in advance of the other, it also often happens that some progress in localization accompanies decomposition. For example, in the preceding sections I often characterized experimental inquiries into operations as inhibiting either the component part or operation. If investigators have also hypothetically linked that operation with a component part, then they can test this by excising the part to inhibit the operation. And likewise investigators can use evidence about operations to confirm that the parts into which they have decomposed the system are indeed working parts. When, therefore, it is possible to coordinate hypotheses about structural and functional decompositions in the course of research, it is advantageous to do so.

### *Testing Models of Mechanisms*

So far in this section I have focused on the discovery of mechanisms. However, science involves not just the advancement of hypotheses but also testing whether they are true in a given situation. While logical empiricists had little to say about discovery, they did attempt to articulate criteria for evaluating proposed laws. Essentially, this involved making predictions from the laws and evaluating the truth of these predictions. The challenge was to articulate a logic that would relate the truth or falsity of a prediction to the confirmation or falsification of the law. This is not the occasion to review the problems and solutions that have been proposed; I simply note the genesis of the problems for confirmation or falsification. Confirmation is challenging because there are always alternative possible laws from which one might make the same prediction (the problem of underdetermination). Falsification is challenging because a false prediction might be due to an error either in the proposed law or in one of the auxiliary hypotheses that figured in deriving the prediction (the problem of credit assignment).

Modulo the difference that scientists typically make predictions from mechanisms by simulating their operation rather than by making logical deduction from laws, the challenge for testing hypotheses is much the same for mechanisms as for laws. A researcher tests hypothesized mechanisms by inferring how the mechanism or its components will behave under specified conditions and uses the results of actually subjecting the system in nature to these conditions to evaluate the proposed mechanism. In principle the same challenges confront tests of mechanisms as tests of laws – (a) different mechanisms might generate the same predictions (underdetermination); and

(b) when a prediction fails, the problem might lie either with the model of the mechanism or with auxiliary hypotheses invoked in making the prediction (credit assignment). Although these problems are not eliminated, they are mitigated when researchers test accounts of mechanisms because their focus is typically not on the mechanism as a whole, but on specific components. Evidence is sought that a given component actually figures in the generation of the phenomenon in the way proposed. Accordingly, predictions are diagnostic – specifically targeted to the effects such a component would have (e.g., that if one intervened in a way known to incapacitate that component, there would be a specific change in the way the mechanism as a whole would behave).<sup>25</sup> If the part does not, for example, perform the operation hypothetically assigned to it, that is often evidence that unsuspected parts are involved and that the operation requires their orchestrated operation.

An important aspect of discovering and testing mechanisms is that inquiry does not simply consist of postulating and testing a mechanism. Rather, research typically begins with an oversimplified account in which only a few components and aspects of their organization are specified. Over time, it is repeatedly revised and filled in (Bechtel & Richardson, 1993). Machamer, Darden, and Craver (2000) refer to the simplified account as a *sketch* of a mechanism. Much of the discovery and testing involved in mechanistic explanation focuses on proposing components or forms of organization that are to be added to or used to revise parts of a sketch, and (often late in the process) localizing the worked-out component operations in the appropriate component part or parts. The entire process is typically a long-term endeavor (Bechtel, 2002a). The result looks much more like a research program (Lakatos, 1970) than like the classical account of theory-testing.

## 8. CONCLUSION

This chapter has focused on providing a general framework within which to understand cell biologists' mission of trying to understand cellular mechanisms. My goal has been to articulate what mechanisms are and how they figure in explanations. Accordingly, I have emphasized the role of diagrams as well as text in representing mechanisms and how simulation underlies reasoning about them. Because mechanisms inherently bridge multiple levels of organization (the components, the mechanism as a whole, and any larger

<sup>25</sup> See, for example, Darden (1990; 1991; 1992) on anomaly resolution in scientific theorizing, a process she compares with the process of redesign of human made artifacts in engineering.

system in which the mechanism functions), I have developed the notion of levels of organization and discussed the respects in which mechanistic explanation is reductionistic. Because organization of biological systems is what distinguishes them from non-biological ones, I also discussed the kinds of organization found in biological mechanisms. I ended with the issue of the discovery and testing of models of mechanisms, a topic that takes us into the main objective of this book – the discovery of mechanisms within the cell. This was a project on which major advances were made in the period 1940–70. As I will discuss in Chapter 4, a major reason these advances occurred when they did was the development of new research techniques – cell fractionation and electron microscopy – that made discovery of cell mechanisms possible. But the project did not emerge from nowhere. In the next chapter I focus on the background prior to 1940 that both provided critical ideas that figured in the later period and also revealed the terra incognita that could not be explored until the new tools were available.